**Dr. Joseph Olive**
**Program Manager**
**Information Processing Technology Office**

# My Mind to Your Mind

Most of us take it for granted that we can communicate easily and clearly with friends, neighbors, and strangers any time we want. But imagine the following: You are standing on an unfamiliar street corner. Suddenly, there's an explosion.

You are confused and frightened, but you can't ask anyone what happened, where to go, or what to do, because no one around you speaks English. There's an announcement on a loudspeaker, but you cannot understand it. You recall some kind of bulletin on TV the night before, but you did not understand that either. Even worse, a few days ago, you found a newspaper with something scribbled on it, but neither you nor your comrades could decipher it.

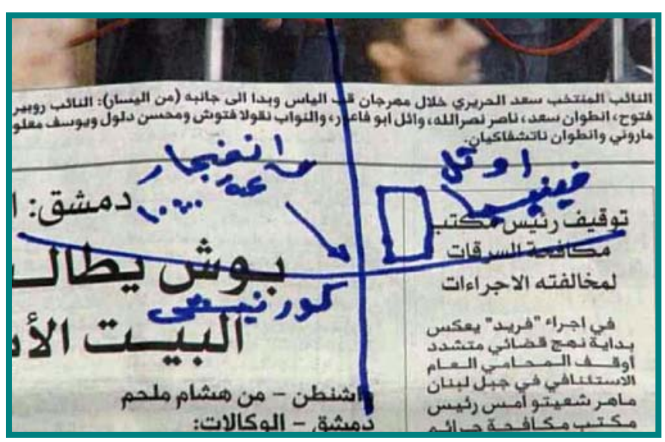
Frightening? Unreal? This is the kind of situation that thousands of our young men and women in the Armed Forces face every day. It clearly illustrates why our Soldiers must be able to communicate with people in foreign lands and understand what is going on around them. It is not possible to send every Soldier through a 63-week course at the Defense Language Institute. And just as impossible to predict where our Soldiers will be deployed next and what language skills they will need.

How can DARPA help? DARPA can develop technology that removes language barriers and connects minds. Technology that makes it possible for our Soldiers to:

- Communicate with allies, enemies, and local populations.
- Use the huge amounts of data available in foreign languages.
- Decipher captured documents.
- Learn foreign languages.

Animals have many ways of communicating: they growl, they lick. Human communication is quite different. Our language is an abstract code that conveys much more information. Sounds form words, and sequences of words form messages.

But, speech is not an exact code. In fact, it can be quite ambiguous. Most languages have many ways of saying the same thing; there are differences among individual voices and speaking styles; and environmental factors, such as noise and echoes, can affect communication.



## My Mind to Your Mind

Humans use a variety of mechanisms to understand spoken messages. We are able to adapt quickly to the sound variations and background noises. We combine knowledge of the language's structure with an understanding of the context of the message.

Here is an example: when I say "Machines may wreck a nice beach," given the context of my talk, you probably thought I said "Machines may recognize speech." You can see how important context is for understanding speech.

What about written language? The first writing systems were pictorial. Languages then developed alphabets related to the speech sounds. While writing is not subject to problems such as degraded sound transmission, it does not convey the complete message, only the words. What is missing in writing are the intonations and timing differences that are so important in spoken communication.

Both spoken and written communications have drawbacks. It is often difficult to hear individual words in an utterance. There are homophones to contend with, words that sound alike but have different meanings. In written text, the equivalent is homographs, words that are spelled alike but have different meanings. Despite these difficulties, humans are very good at transferring information from one mind to another.

What about communicating with computers?

In "2001: A Space Odyssey," HAL had complete language facility. But that was—and still is—science fiction. In reality, computers do not have the knowledge, experience, deductive reasoning, or language comprehension needed to understand speech and text. Fortunately, DARPA has long understood the importance of human-computer communication and works hard to meet the military's need. Lives—often many lives—can be on the line.

One area of interest is computerized speech-to-speech translation. Another is developing automated means for digesting the enormous volumes of information generated in multiple languages all over the world. In the Arab speaking world, there are scores of radio stations, television stations, newswires, and web sites producing hundreds of hours of audio and megabytes of text

every day. These information streams must be transcribed, translated, and distilled to find the tiny scraps relevant to us, and it must be done in almost real time. (A delay of just a few hours may be too long.) Translating all this information manually requires more skilled people than are available.

DARPA's solution is to develop advanced language technologies, and for 3 decades, we have been DoD's primary supporter of research in this area. In the beginning, our research was restricted to very basic word recognition because of our limited understanding of the science involved, computer speed, and storage capabilities. Recognition algorithms were trained for individual users, making the systems speaker dependent. Progress was very slow. So slow that, in 1969, John Pierce, head of human sciences research at Bell Labs, wrote:

> General purpose speech recognition seems far away. Special purpose speech recognition is severely limited. It would seem appropriate for people to ask themselves why they are working in the field and what they expect to accomplish.

Fortunately, persistence paid off, and DARPA produced new theories and methods. Machine vocabularies increased and speech recognition systems became speaker independent. DARPA developed applications for spotting key words and concepts in natural speech, a technology that migrated to the commercial sector and can be found today in call centers and travel reservation services. Eventually, DARPA developed systems using unconstrained vocabularies that today are the basis of speech recognition systems, such as IBM's ViaVoice and Dragon Naturally Speaking. Simultaneously, DARPA supported research in text processing, especially in the areas of language translation and information retrieval.

Given that computers are unable to understand the language, how have we managed to get so far? We did it by treating language as stochastic processes. For speech, we train models for each sound of the language, assuming some probability distribution. We compute the probabilities of word sequences from huge text databases. To recognize a new utterance, we compute the probabilities of sound sequences. Using a dictionary, we eliminate sound sequences that cannot occur in the language. We construct a network of all possible sound and word sequences with their associated probabilities, and the most likely path through the network yields the recognized utterance. Similar stochastic models are used for text processing.

DARPA's long-term investment in language has produced technologies with extremely valuable military applications. One is a one-way translation device, Phraselator, used today in Iraq and Afghanistan. The Phraselator is designed for very specific applications, such as working in a medical environment, so its language repertoire is quite constrained. Here's what our troops have to say about it:

> … And it's been an invaluable tool. Not only to our command. But I mean, you know, I have five kids I want to go home to, and this might help.

> ….through an interpreter. You don't know his background. So if you have secure information you are trying to pass, or to obtain. If you can do it by yourself with a machine, then you don't have to worry about any type of a compromising situation.

A new effort is moving toward two-way translation, again, in limited domain applications. The goal is for Service personnel to talk in natural English and have their speech translated automatically into Arabic, and to hear spoken Arabic responses translated into English.

Another byproduct of DARPA's research is a system that monitors open sources, TV and radio transmissions, and news articles on the web.

The system is able to transcribe and translate Arabic news in real time.

This system is far from perfect. But, it is good enough to extract basic information from an article or broadcast and allow analysts to decide whether they'd like it translated more thoroughly by humans. Since we do not have the manpower to translate every Arabic news source, this system is a valuable way to reduce the flood of Arabic media down to a manageable volume.

Is it effective? Well, CENTCOM receives about 5,000 Arabic language articles per week. After translating and filtering these automatically, CENTCOM analysts reduce that number to about 300 that are important enough to forward to human translators; this is a 16:1 reduction. This system has proven invaluable to our fighting forces. US

## My Mind to Your Mind

Forces received information that insurgents were hiding inside a mosque. Unfortunately, no one had an image of the mosque. Using the system, an image was found in an Al Jazeera video clip and our Soldiers soon had the photo they needed.

Such examples demonstrate that DARPA has come a long way in processing speech and text, but we still have far to go. We have just begun a very aggressive program, GALE, that is developing technology to do the following quickly and accurately:

- Transcribe speech into English text, regardless of the source language.
- Translate foreign text into English text.
- Identify and filter the precise, actionable information needed by commanders.

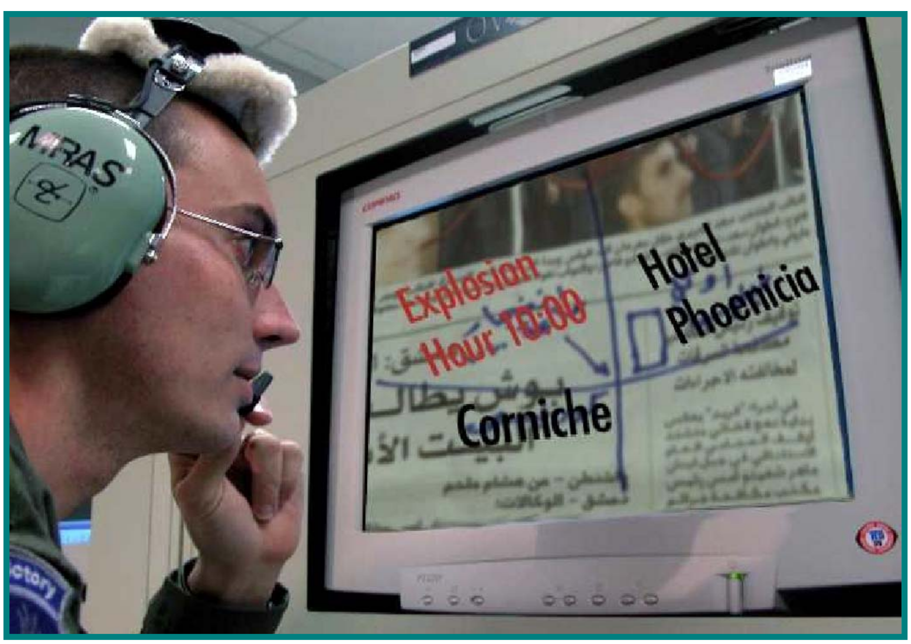
This is a technology program that is truly DARPA-hard.

We fully expect to succeed through vigorous, multidisciplinary research, bringing together previously separate efforts and developing tightly coupled solutions. For example, we can use topic identification to help transcription and translation. We can use accurate name tagging to avoid trying to translate names. We can use accurate parsing to prevent the computer from producing sentences like


"which classified the United States as a sponsor of terrorism" when the Arabic author actually wrote "which is classified by the United States as a sponsor of terrorism." The machine confused the subject and object.

These are DARPA's existing programs. What about the future?

Our plans for the future include attacking challenges such as unconstrained speech-to-speech and text-to-text translation, automatic conversion and translation of captured documents, including poor quality copies, documents that mix printing and handwriting, and documents written in more than one language. We're also focusing on ways to process speech in noisy and echoic environments.




We have big challenges and we need your help. More important, our men and women in uniform need your help, so that one day soon the scenario we described, instead of ending in tragedy, will go more like: the scrap of paper the Soldier found containing a map of the location of the bomb is read automatically and transmitted to headquarters, warning all personnel in the vicinity to stay away from the designated corner.

Mission accomplished!